FermiCloud

Enabling Scientific Computing with Integrated Private Cloud Infrastructures

Steven Timm
Fermilab Grid and Cloud Computing Dept.
for FermiCloud Project
K. Chadwick, F. Lowe, D. Yocum, T. Hesselroth, G. Garzoglio, P. Mhashilkar, T. Levshina

SAUCG workshop
July 18, 2011

Work supported by the US Dept. of Energy under contract DE-AC02-07CH11359

What FermiCloud is

- Infrastructure-as-a-service private cloud for Fermilab Scientific Program.
- Integrated into Fermilab site security structure.
- Virtual machines have full access to existing ***Fermilab network*** and mass storage devices.
- Scientific stakeholders get ***on-demand access*** to virtual machines without sysadmin intervention.
- Virtual machines created by users and destroyed or suspended when no longer needed.
- Testbed for developers and integrators to evaluate new grid and storage applications on behalf of scientific stakeholders.
- Ongoing project to build and expand the facility
  - Technology evaluation, requirements, deployment.
  - Scalability, monitoring, performance improvement.
  - High availability and reliability.

Virtualization at Fermilab

- FermiGrid Services
  - Highly Available provisioned virtual services
  - SLF5+Xen
- General Physics Compute Facility
  - Deployment of experiment-specific virtual machines for Intensity Frontier experiments
  - Oracle VM (Commercialized Xen)
- Virtual Services Group
  - Virtualization of Fermilab business systems using VMWare
  - Windows

Science Facility Drivers for FermiCloud

- Continue program of virtualizing all scientific servers that can be virtualized. Many experiment servers need minimal CPU, memory but want ports to themselves.
- Improve utilization of power, cooling and employee time (admins and developers) for managing small science servers.
- Had to replace 6 racks of legacy development machines with limited hardware budget and computer room space.
- CERN IT + HEPiX Virtualisation Taskforce program to have uniformly-deployable virtual machines. Expect LHC and future Fermilab experiments will eventually require cloud technology.

Science Stakeholders

- Joint Dark Energy Mission->WFIRST->LSST
  - Distributed messaging system, testing fault tolerance.
- Fermilab Intensity Frontier
  - Monitoring Server (MCAS)
  - GridFTP endpoint server
  - Experiment-specific storage investigations
- Fermilab D0 Experiment
  - Job Forwarding Server
- Extenci project (Cloud activities, LHC)
  - Distributed storage on WAN.
- GEANT4
  - Validation server
- Scientific middleware development hosting
  - Host developers and integrators of OSG middleware.

Hardware

- 2x Quad Core Intel Xeon E5640 CPU
- 2 SAS 15K RPM system disk 300GB
- 6x 2TB SATA disk
- LSI 1078 RAID controller
- Infiniband card
- 24GB RAM
- 23 machines total
- Arrived June 2010
- +25TB Bluearc NAS disk

Software Technologies

- OS: Scientific Linux 5, 6
- Hypervisor: KVM
- Cloud Management: OpenNebula
- Modifications to OpenNebula CLI, Query API, GUI to use X.509 authentication to launch virtual machines. (See my talk at OGF32.)
- Secure credential store—all security secrets loaded at runtime only.
- Site-wide patching and vulnerability scanning facilities.

Current Technology Investigations

- Testing storage services with real neutrino experiment codes, identify NFS alternatives.
- Using Infiniband interface to create sandbox for MPI applications.
- Batch queue look-ahead to create worker node VM's on demand.
- Submission of multiple worker node VM's, grid cluster in the cloud.
- Idle VM detection and suspension, backfill with worker node VM's.
- Leverage site “network jail” for new virtual machines.
- IPv6 support.

High Availability and Service Levels

- Add SAN for live migration and large datablock capacity.
- Split FermiCloud between two buildings
- Mirror storage between two buildings
- Set up high-availability procedures for failover of cloud controller and migration of virtual machines.
- Offer three service levels
  - High availability 24x7
  - Regular virtual machine
  - Opportunistic (spot market) can be pre-empted anytime.
- Stakeholders billed for usage according to an economic model, analogous to existing tape robot facility.

Ongoing software development

- Accounting and billing—Cloud accounting add-ons to Gratia accounting project.
- Monitoring—How many machines are running, who is running them, is everything up that should be up?
- Authorization—Apply well-tested and interoperable grid authorization tools to cloud authorization as well.
- All of above in collaboration with other projects and standards bodies.

Using Virtualization to Enable Science

- New interactive science applications that require ongoing interaction or unique network topologies and don't fit grid batch processing paradigm.
- Complicated software stacks where grid distribution has been difficult or impossible.
- Legacy experiments which require specific OS and library combinations.
- Extra compute capacity on demand for experiments that need it.
- Virtualization used on 32-core+ worker nodes to
  - Pin applications to appropriate CPU-memory combinations for better performance
  - Sandbox applications to keep one rogue job from killing the other 31.
  - Memory segments can grow or be shared as needed.

Conclusions

- FermiCloud has successfully deployed a wide range of servers for the scientific program.
- FermiCloud has been a testbed for several evaluations of storage and middleware that benefit the scientific program.
- FermiCloud has already provided significant power and cooling savings, and significant convenience benefits to scientific stakeholders
- Now integrating our work with other internal Fermilab virtualization activities and external projects.
- We welcome interest from new users, stakeholders, and other cloud-based projects.